REMARKS

Claims 1, 5-11 and 36-48 were pending. Claims 1, and 6-11 were allowed. Claims 5 and 36-48 have been rejected. Claims 46-48 have been amended. New claim 49 has been added.

Support for the amendments to the claims and for new claim 49 can be found throughout the application as originally filed including, for example, on page 12, lines 12-16.

No new matter has been added.

Upon entry of this paper, claims 1, 5-11 and 36-49 will be pending.

Withdrawn Objections/Rejections

The Office noted that the following objections/rejections were withdrawn:

(a) objection of claims 39, 45 and 46; and

(b) rejection of claims 45-47 under 35 U.S.C. §112, second paragraph.

Claim Objections

Claim 46 was objected to due to the repeated word "the". Applicants have revised claim 46 to remove this obvious typographical error.

Claim Rejections under 35 U.S.C. § 112, first paragraph

New matter

Claim 47 was rejected under 35 U.S.C. § 112, first paragraph, as allegedly containing new matter. Specifically the Office alleges that there is no support for the limitations "46-1173 of SEQ ID NO:1" As suggested by the Examiner, Applicants have amended claim 47 to correct the listed nucleotide range from "46-1173" to "446-1173". As pointed out by the Office, support for this revision can be found, for example, on page 22 of the specification.

Claim 48 was rejected under 35 U.S.C. § 112, first paragraph, as allegedly containing new matter. Specifically the Office alleges that there is no support for the:

> isolated nucleic acid molecule comprising a polynucleotide encoding a polypeptide; a polynucleotide at least 95% identical to a polynucleotide; an a polynucleotide encoding a polypeptide at

> least 95% identical to SEQ ID NO:2 wherein the nucleic acid molecule comprises nucleotides 365-1173 of SEQ ID NO:1

further stating that:

> the context of this recitation does not overlap with the claimed invention. In essence, this passage within the specification does not support the breadth of the claims and further more it seems nucleotides 365-1173 of SEQ ID NO:1 do not encode amino acids 1 to 273; amino acids 2 to 273; nor 26 to 273 of SEQ ID NO:2.

Applicants respectfully disagree.

Claim 48 is directed to a subset of the nucleic acids molecules set forth in claims 5, 36 and 39, each of which recites nucleic acids comprising a polynucleotide which encodes a polypeptide related to SEQ ID NO:2. For example, claim 5 provides a nucleic acid molecule comprising a polynucleotide which encodes a polypeptide comprising amino acids 1-273, 2-273, or 26-273 of SEQ ID NO:2, with up to 5 conservative amino acid substitutions as compared to SEQ ID NO:2. Claim 36 provides a nucleic acid molecule comprises a polynucleotide at least 95% identical to a polynucleotide which encodes a polypeptide comprising amino acids 1-273, 2-273, or 26-273 of SEQ ID NO:2. Claim 39 recites a nucleic acid molecule comprises a polynucleotide which encodes a polypeptide at least 95% identical to the polypeptide of SEQ ID NO:2. Applicants further note that each of claims 5, 36 and 39 recite that the encoded polypeptide is expressed at a higher level in metastatic cells relative to non-metastatic cells.

Each of claims 5, 36 and 39 provide some variability in the nucleotide sequence. In claim 5, the variability is provided through the recitation that the encoded polypeptide may have up to 5 conservative substitutions as compared to the reference sequence, SEQ ID NO:2. Claim 36 provides variability in the nucleotide sequence through the recitation that the polynucleotide is at least 95% identical to the polynucleotide encoding the amino acid of SEQ ID NO:2. Claim 39 allows variability in the nucleotide sequence through its recitation that the polynucleotide encodes a polypeptide at least 95% identical to the amino acid of SEQ ID NO:2.

Claim 48 provides further identifying characteristics of the claimed polynucleotide, specifically reciting that the portion of the polynucleotide which encodes the amino acid of SEQ ID NO:2 (or 95% homolog thereof) comprises nucleotides 365-1173 of SEQ ID NO:1. In other

words, in claim 48, the variability in the nucleotide sequence allowed through claims 5, 36 and 39 does *not* occur at the region corresponding to nucleotides 365-1173 of SEQ ID NO:1.

The Office points out correctly that nucleotides 365-1173 of SEQ ID NO:1 do not encode amino acids 1-273, 2-273, or 26-273 of SEQ ID NO:2. Instead, nucleotides 365-1173 of SEQ ID NO:1 represents a *portion* of a nucleotide sequence which encodes amino acids 1, 2, or 26 to 273 of SEQ ID NO:2. As set forth above and as depicted in the attached page, nucleotides 365-1173 represent a region of SEQ ID NO:1173 which is conserved in some embodiments. Applicants note that the open reading frame (orf) encoding hsOAF (SEQ ID NO:2) spans nucleotides 356 to 1176 of SEQ ID NO:1. Nucleotides 365-1173 of SEQ ID NO:1 fall within this orf.

In view of the foregoing Applicants respectfully request the withdrawal of the new matter rejection.

Written description

Claims 5 and 36-48 stand rejected under 35 U.S.C. § 112, first paragraph, as allegedly failing to comply with the written description requirement. Although acknowledging that the specification "sets forth wild type hsOAF (polynucleotide, SEQ ID NO:1, which encodes polypeptide, SEQ ID NO:2 in its entirety)", the Office alleges that the "written description is not commensurate in scope with claims drawn to variants of SEQ ID NO:1 and SEQ ID NO:2, which have not been defined by functional or structural characteristics'." (Office Action, page 5). For at least the reasons that follow, Applicants respectfully disagree.

Applicants respectfully assert that characteristics of the claimed sequences have been provided which would "allow persons of ordinary skill to recognize that [he or she] invented what is claimed" As recognized by the Office, the court in *Regents if the University of California v. Eli Lilly* noted that an adequate written description of DNA "requires a precise definition, such as by structure, formula, chemical name or physical properties ..." The pending claims conform with the guidance provided by the Federal Circuit. Each of claims 5, 36 and 39 provide a sequence of a representative species. Each of claims 5, 36 and 39 further disclose a required physical property of the claimed sequences, that is, each recites that members of the

claimed genus include a polynucleotide that encodes a polypeptide which is expressed at a higher level in metastatic cells relative to non-metastatic cells.

In *Ex parte Sun* (Appeal No. 2003-1993; copy enclosed), the Board of Patent Appeals and Interferences considered the appropriateness of rejections under the written description and enablement requirements where the specification discloses a molecule within the claims and a functional assay for activity. The Board explained in *Ex parte Sun* that the following claim was illustrative of those on appeal.

> 31. An isolated wee1 nucleic acid molecule comprising a member selected from the group consisting of:
>
> (a) a polynucleotide that encodes a polypeptide of SEQ ID NO:2;
>
> (b) a wee1 polynucleotide having at least 80% identity to the entire coding region of SEQ ID NO:1;
>
> (c) a polynucleotide comprising the coding sequence set forth in SEQ ID NO:1; and
>
> (d) a polynucleotide complementary to a polynucleotide of (a) through (c).

The specification of the application on appeal disclosed that SEQ ID NO:2 encoded a protein having a defined function (similar to that of a known tyrosine kinase). The specification explained that the protein is useful in genetic engineering of corn plants to increase productivity. The examiner had rejected claim 31 as failing to meet the written description requirement, arguing that one skilled in the art could not predict the structure and function of nucleic acid "comprising a wee1 polynucleotide having at least 80% identity to the entire coding region of SEQ ID NO:1" The examiner had also argued that the specification did not "teach a single representative species with 80% identity and WEE1 function".

After reviewing the relevant case law, including *Lilly* and *Enzo Biochem, Inc. v. Gen-Probe Inc.*, 296 F.3d 1316 (Fed. Cir. 2002), the Board concluded that the rejected claims, including claim 31, met the written description requirement. The Board pointed out that the specification describes the sequence of a nucleic acid molecule encoding SEQ ID NO:2 and the sequence of a nucleic acid molecule comprising the coding sequence of SEQ ID NO:1. The Board also noted that the specification provides a description how to screen for WEE1 activity. The Board concluded that,

> [I]t would reasonably appear that such a description in the specification would constitute sufficiently detailed, relevant identifying characteristics of the claimed subject matter consistent with *Enzo*.
> In our view, the examiner has failed to indicate why one of ordinary skill in the art, who is in possession of the very specific chemical structures of a polynucleotide that encodes a polypeptide of SEQ ID NO:2 and a polynucleotide comprising the coding sequence set forth in SEQ ID NO:1, would be unable to recognize, upon reading the disclosure, that appellants invented the claimed subject matter, including homologues sharing structural features with the specifically claimed and disclosed structures.

In the present case, as in *Ex parte Sun*, a functional limitation is provided. One of skill in the art at the time the present application was filed would be able to determine whether sequences falling within the claimed genus satisfied the functional limitation, e.g. whether encoded polypeptides are expressed at higher levels in metastatic cells relative to non-metastatic cells. In *Ex parte Sun* the Board further concluded that the written description requirement was met even though only a single species is disclosed.

Accordingly, it is clear that the disclosure of even a *single* species combined with a functional assay provides an adequate written description for a claim to a genus of molecules. A person skilled in the art would recognize, upon reading the disclosure, that Applicants invented the claimed subject matter, including 95% homologs or homologs with up to 5 conservative amino acid substitutions as compared to SEQ ID NO:2, sharing structural features with the specifically claimed and disclosed sequences.

Applicants further note that the written description for the pending claims provided in the present specification conforms with the guidelines provided by the Patent Office's guidelines on the subject: Synopsis of Application of Written Description Guidelines, www.uspto.gov/web/menu/written.pdf ("Guidelines"). Example 14 of the Guidelines illustrates a hypothetical situation that mirrors the present case. Example 14 provides an example of a product by function claim, where the specification teaches that SEQ ID NO:3 and "variants that are at least 95% identical to SEQ ID NO:3 and catalyze the reaction of A→B" are essential to the operation of the claimed invention.

Example 14 then provides the following guidance to examiners:

> The specification indicates that the genus of [nucleic acids] that must be variants of SEQ ID NO:3 does not have substantial

> variation since all of the variants must possess the specified catalytic activity and must have at least 95% identity to the reference sequence, SEQ ID NO:3. The single species disclosed is representative of the genus because all members have at least 95% structural identity with the reference compound and because of the presence of an assay which applicant provided for identifying all of the at least 95% identical variants of SEQ ID NO:3 which are capable of the specified [kinase-encoding] activity. One of skill in the art would conclude that applicant was in possession of the necessary common attributes possessed by the members of the genus.
> Conclusion: The disclosure meets the requirements of 35 USC §112 first paragraph as providing adequate written description for the claimed invention.

As in the Example 14 hypothetical, the present claims are drawn to a genus of molecules whose "variants must possess the specified activity and must have identity to" a reference sequence. The present application discloses the species SEQ ID NO:1 and information relating to variants. Applicants respectfully submit that new claim 49 is also supported by adequate written description and is consistent with Example 14 of the Guidelines.

Therefore, because one of skill in the art would recognize that Applicants were in possession of the claimed invention at the time the application was filed, Applicants respectfully request withdrawal of this rejection of the claims under 35 U.S.C. §112, first paragraph.

Enablement

Claims 5 and 36-48 stand rejected under 35 U.S.C.§112, first paragraph, as allegedly failing to comply with the enablement requirement. The Office alleges that the specification fails to enable "variants that have at least 95% sequence identity to polynucleotides that encode SEQ ID NO:2." The Office further alleges that the "nucleic acids [which are 95% identical to SEQ ID NO:1] will encode proteins that may not maintain the activities proposed in the specification, such as a marker for distinguishing between tumors, which will or have metastasized." Also, the Office asserts that "predictability of which changes can be tolerated in a polypeptide's amino acid sequence and still retain similar activity requires a knowledge and guidance with regard to which amino acid or acids in the polypeptide's sequence, if any, are tolerant of modification and which are conserved and detailed knowledge of the ways in which

the protein's structure relates to its function." (Office Action, pages 8-9). The Office also cites Lazar et al. (Molecular and Cellular Biology 8(3):1247-1252) as evidence that "the introduction of mutations in an amino acid sequence will yield products with different biological activity from the wild type protein." (Office Action, page 9). Because the specification adequately enables the claimed invention, Applicants respectfully disagree.

Preliminarily, with respect to the Office's comment that nucleotide sequences with less than 100% identity to SEQ ID NO:1 "may not maintain the activities proposed in the specification ... such as a marker for distinguishing between tumors ...", Applicants note that claim 5, for example, requires that "the encoded polypeptide is expressed at a higher level in metastatic cells relative to non-metastatic cells", the same activity as set forth for SEQ ID NO:2, the amino acid encoded for by a nucleotide sequence of SEQ ID NO:1. Applicants further note that Figure 7 of the application as originally filed provided a graphical representation of residues conserved between human and Drosophila OAF amino acid sequences. Regions with dark shading indicated conserved residues while light shading indicates conservative substitutions. Accordingly, based, for example on the data set forth in Figure 7, the skilled artisan could readily predict which residues "are tolerant of modification" and which residues "are conserved".

Claim 5 and new claim 49 are also in compliance with the enablement requirement. Based, for example, on Figure 7, one of skill in the art could readily identify "no more than 5/3 conservative amino acid substitutions" and test such sequences to determine if the encoded polypeptide is expressed at a higher level in metastatic cells relative to non-metastatic cells without any undue experimentation.

With respect to the Lazar reference, Applicants note that Lazar is *not* related to the sequences presently claimed. Lazar instead pertains to transforming growth factor alpha. Although Applicants agree that it is possible, at least in some cases, to abolish activity of a given protein by mutating a single critical residue, Applicants disagree that this fact means that one of ordinary skill cannot make functional analogs of the claimed proteins without undue experimentation. In fact, Applicants point out that several references teach that substitutions of single residues are not as "critical" as suggested by the Office. Indeed, Lazar even notes that of the seven transforming growth factor alpha mutants generated, only one exhibited a complete loss of activity, while two mutants retained biological activity, two mutants exhibited reduced

biological activity, and two mutants showed low levels of activity). The Bowie reference (Bowie et al. (1990) Science 247:1306-10; copy attached hereto) confirms the resilience of proteins of such substitutions, stating that "proteins are surprisingly tolerant of amino acid substitutions." (*see,* Bowie et al., at page 1306, col.2, lines 12-13). Bowie et al. cites as evidence a study carried out on the lac repressor. Of approximately 1500 single amino acid substitutions at 142 positions in this protein, about one-half of the substitutions were found to be "phenotypically silent": that is, had ***no*** noticeable effect on the activity of the protein (page 1306, col. 2, lines 14-17). Presumably the other half of the substitutions exhibited effects ranging from slight to complete abolishment of repressor activity. Thus, based on Bowie et al.'s teachings, one can expect to find over half (and very possibly well over half) of random substitutions in any given protein to result in mutated proteins with full or nearly full activity. These are far better odds than those at issue in *In re Wands*, 858 F.2d 731 (Fed. Cir. 1988), in which the court said that screening many hybridomas to find the few that fell within the claims was not undue experimentation. The question is not whether it is possible to abolish activity with a point mutation (as the Office seems to suggest), but rather whether one of ordinary skill can produce, without undue experimentation, mutants in which the activity is not abolished.

Accordingly, based on the disclosure of Bowie et al., one would predict that even random substitution of residues in the presently claimed proteins will predictably result in a majority of the variants having full or partial activity. Given the information provided in the specification regarding conserved residues in hOAF (see, for example Fig. 7), one of ordinary skill would know to avoid those residues or make only conservative changes there, thereby making the predictability of success even higher than in the lac repressor study. Furthermore, the specification amply teaches how to make and test mutants to find those with the expression pattern required by the claims. The Office's view that "the predictability of changes to the amino acid sequence is practically nil as far as biological activities are concerned" does not appear to be warranted in view either of Lazar et al. or in view of the specification or other references available at the time the present application was filed.

The specification describes several assays that can be readily carried out by those of ordinary skill in the art with any variant polypeptide having the requisite degree of identity or

with the recited number of conservative amino acid substitutions (see, for example, page 14 of the specification which provides procedures to identify amino acids essential for function). The specification further provides assays that can be used to confirm activity of polypeptides (see, for example, the proliferation and invasion assay described in the specification at page 36). Manipulating the sequence of a nucleic acid requires only routine techniques, and those of ordinary skill in the art were well able to generate, analyze, and use variant nucleic acids and variant polypeptides at the time the present application was filed.

In view of these considerations, Applicants request withdrawal the rejections for lack of enablement.

Applicants respectfully request the withdrawal of the rejections under 35 U.S.C.§112, first paragraph.

Allowed Claims

Applicants thank the Examiner for the indication that claims 1 and 6-11 are allowed.

Related Applications

Applicants call the Examiner's attention to the following related applications:

10/200,026; filed July 18, 2002, now US Patent 7,081,517; and

11/450,251 filed on June 8, 2006, pending.

These applications are available in PAIR and the Examiner is encouraged to review them.

CONCLUSION

The foregoing represents a *bona fide* attempt to advance the present application to allowance. Applicants respectfully assert that all claims are in condition for allowance, which action is hereby requested. The Examiner is invited to telephone the undersigned attorney at (302) 778-8458 if such would expedite prosecution.

Please apply any charges or credits to deposit account 06-1050.

Respectfully submitted,

Date: **November 30, 2007**

Gwilym J.O. Attwell
Reg. No. 45,449

Fish & Richardson P.C.
P.O. Box 1022
Minneapolis, MN 55440-1022
(302) 652-5070 telephone
(877) 769-7945 facsimile

Open Reading Frame (ORF) of SEQ ID NO:1 and encoded protein (SEQ ID NO:2)

```
 356 atgcgccttcccggggtacccctggcgcgccctgcgctgctgctg
     M  R  L  P  G  V  P  L  A  R  P  A  L  L  L
 401 ctgctgccgctgctcgcgccgctgctgggaacgggtgcgccggcc
     L  L  P  L  L  A  P  L  L  G  T  G  A  P  A
 446 gagctgcgggtccgcgtgcggctgccggacggccaggtgaccgag
     E  L  R  V  R  V  R  L  P  D  G  Q  V  T  E
 491 gagagcctgcaggcggacagcgacgcggacagcatcagcctcgag
     E  S  L  Q  A  D  S  D  A  D  S  I  S  L  E
 536 ctgcgcaagcccgacggcaccctcgtctccttcaccgccgacttc
     L  R  K  P  D  G  T  L  V  S  F  T  A  D  F
 581 aagaaggatgtgaaggtcttccgggccctgatcctgggggagctg
     K  K  D  V  K  V  F  R  A  L  I  L  G  E  L
 626 gagaaggggcagagtcagttccaggccctctgctttgtcacccag
     E  K  G  Q  S  Q  F  Q  A  L  C  F  V  T  Q
 671 ctgcagcacaatgagatcatccccagtgaggccatggccaagctc
     L  Q  H  N  E  I  I  P  S  E  A  M  A  K  L
 716 cggcagaaaaatccccgggcagtgcggcaggcggaggaggttcgg
     R  Q  K  N  P  R  A  V  R  Q  A  E  E  V  R
 761 ggtctggagcatctgcacatggatgtcgctgtcaacttcagccag
     G  L  E  H  L  H  M  D  V  A  V  N  F  S  Q
 806 ggggccctgctgagcccccatctccacaacgtgtgtgccgaggcc
     G  A  L  L  S  P  H  L  H  N  V  C  A  E  A
 851 gtggatgccatctacacccgccaggaggatgtccggttctggctg
     V  D  A  I  Y  T  R  Q  E  D  V  R  F  W  L
 896 gagcaaggtgtggacagttctgtgttcgaggctctgcccaaggcc
     E  Q  G  V  D  S  S  V  F  E  A  L  P  K  A
 941 tcagagcaggcggagctgcctcgctgcaggcaggtgggggaccgc
     S  E  Q  A  E  L  P  R  C  R  Q  V  G  D  R
 986 gggaagccctgcgtctgccactatggcctgagcctggcctggtac
     G  K  P  C  V  C  H  Y  G  L  S  L  A  W  Y
1031 ccctgcatgctcaagtactgccacagccgcgaccggcccacgccc
     P  C  M  L  K  Y  C  H  S  R  D  R  P  T  P
1076 tacaagtgtggcatccgcagctgccagaagagctacagcttcgac
     Y  K  C  G  I  R  S  C  Q  K  S  Y  S  F  D
1121 ttctacgtgccccagaggcagctgtgtctctgggatgaggatccc
     F  Y  V  P  Q  R  Q  L  C  L  W  D  E  D  P
1166 tacccaggctag     1177
     Y  P  G  *
```

Paper No. 27

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE BOARD OF PATENT APPEALS AND INTERFERENCES

Ex parte YUEJIN SUN, BRIAN R. DILKES, BRIAN A. LARKINS, KEITH S. LOWE, WILLIAM J. GORDON-KAMM and RICARDO A. DANTE

Appeal No. 2003-1993
Application No. 09/470,526

ON BRIEF

Before WILLIAM F. SMITH, MILLS and GRIMES, Administrative Patent Judges.

MILLS, Administrative Patent Judge.

DECISION ON APPEAL

This is a decision on appeal under 35 U.S.C. §134 from the examiner's final rejection of claims 2-11, 31, 33 and 35-36 which are the claims on appeal in this application. Claims 14, 32 and 37 have been allowed.

Claim 31 is illustrative of the claims on appeal and reads as follows:

31. An isolated wee1 nucleic acid comprising a member selected from the group consisting of:
(a) a polynucleotide that encodes a polypeptide of SEQ ID NO:2.;
(b) a wee1 polynucleotide having at least 80% identity to the entire coding region of SEQ ID NO:1;

(c) a polynucleotide comprising the coding sequence set forth in SEQ ID NO:1; and
(d) a polynucleotide complementary to a polynucleotide of (a) through (c).

The prior art references relied upon by the examiner are:

Aligue et al. (Aligue), "Regulation of *Schizosaccharomyces pombe* Wee1 Tyrosine Kinase," J. Biol. Chem., Vol. 272, pp. 13320-13325 (1997)

Hemerly et al. (Hemerly), "Dominant negative mutants of the Cdc2 kinase uncouple cell division from iterative plant development," The EMBO Journal, Vol. 14, pp. 3925-3936 (1995)

Grounds of Rejection

Claims 2-11, 31, 33 and 35-36 stand rejected under 35 U.S.C. § 112, first paragraph as containing subject matter that was not described in the specification in such a way as to reasonably convey to one skilled in the art at the time the application was filed that the inventor had possession of the claimed invention.

Claims 2-11, 31, 33 and 35-36 stand rejected under 35 U.S.C. § 112, first paragraph for lack of enablement.

These rejections are reversed.

DISCUSSION

In reaching our decision in this appeal, we have given consideration to the appellants' specification and claims, to the applied references, and to the respective positions articulated by the appellants and the examiner.

Rather than reiterate the conflicting viewpoints advanced by the examiner and the appellants regarding the noted rejections, we make reference to the examiner's Answer for the examiner's reasoning in support of the rejection, and to the appellants' Brief for the appellants' arguments thereagainst. As a consequence of our review, we make the determinations which follow.

Background

The subject matter of the present application is generally directed to corn plant nucleic acids and their encoded proteins which are involved in cell cycle regulation. Specification, page 4. In particular, the claimed invention is directed to a wee1 homologue from maize, zmwee1, whose activity resembles related protein tyrosine kinases. Specification, page 6. The zmwee1 protein is indicated in the specification to be useful in the genetic engineering of the corn plant to increase maize productivity. Specification, page 3.

More specifically, claim 31 is directed to an isolated wee1 nucleic acid comprising a member selected from the group consisting of: a polynucleotide that encodes a polypeptide of SEQ ID NO:2.; a wee1 polynucleotide having at least 80% identity to the entire coding region of SEQ ID NO:1; a polynucleotide comprising the

coding sequence set forth in SEQ ID NO:1; and a polynucleotide complementary to a polynucleotide described above.

According to the prior art, Aligue, Wee1 tyrosine kinase regulates mitosis by carrying out the inhibitory tyrosine 15 phosphorylation of Cdc2 M-phase inducing kinase. Abstract. The specification confirms this, stating "induced wee1 overexpression results in phosphorylation of p34 at tyrosine-15 (inactivating p34), effectively blocking the transition from G2 into mitosis." Specification, page 37. The "encoded [wee1] protein is an important part of the checkpoint control machinery that regulates $p34^{cdc2}$ activity and it's [sic] participation in the active MPF (maturation promoting factor) complex." Specification, page 36. Wee1 activity can be stimulated by the CDK2-cyclin A complex, or inhibited by nim1. Specification, page 36.

Description

Claims 2-11, 31, 33 and 35-36 stand rejected under 35 U.S.C. § 112, first paragraph as containing subject matter that was not described in the specification in such a way as to reasonably convey to one skilled in the art at the time the application was filed that the inventor had possession of the claimed invention.

The Federal Circuit has discussed the application of the written description requirement of the first paragraph of § 112 to inventions in the field of biotechnology. See University of California v. Eli Lilly and Co., 119 F.3d 1559, 1568, 43 USPQ2d 1398, 1406 (Fed. Cir. 1997). The court explained that

> In claims involving chemical materials, generic formulae usually indicate with specificity what the generic claims encompass. One skilled in the art can distinguish such a formula from others and can identify many of the species that the claims encompass. Accordingly, such a formula is normally an adequate description of the claimed genus . . . [H]owever, a generic statement such as "vertebrate insulin cDNA" or "mammalian insulin cDNA," without more, is not an adequate written description of the genus because it does not distinguish the claimed genus from others, except by function. It does not specifically define any of the genes that fall within its definition. It does not define any structural features commonly possessed by members of the genus that distinguish them from others. One skilled in the art therefore cannot, as one can do with a fully described genus, visualize or recognize the identity of the members of the genus. A definition by function, as we have previously indicated, does not suffice to define the genus because it is only an indication of what the gene does, rather than what it is.

Id.

The Lilly court also stated that "[a] written description of an invention involving a chemical genus, like a description of a chemical species, 'requires a precise definition, such as by structure, formula, [or] chemical name,' of the claimed subject matter sufficient to distinguish it from other materials." Id. at 1567, 43 USPQ2d at 1405. Finally, the court addressed the manner by which a genus of cDNAs might be described. "A description of a genus of cDNAs may be achieved by means of a recitation of a representative number of cDNAs, defined by nucleotide sequence, falling within the scope of the genus or of a recitation of structural features common to the members of the genus, which features constitute a substantial portion of the genus." Id. at 1568, 43 USPQ2d at 1406.

The Federal Circuit has also addressed the written description requirement in the context of DNA-related inventions. See Enzo Biochem, Inc. v. Gen-Probe Inc., 296 F.3d 1316, 63 USPQ2d 1609 (Fed. Cir. 2002). The Enzo court adopted the standard that "the written description requirement can be met by 'showing that an invention is complete by disclosure of **sufficiently detailed, relevant identifying characteristics** . . . i.e., complete or partial structure, other physical and/or chemical properties, functional characteristics when coupled with a known or disclosed correlation between function and structure, or some combination of such characteristics.'" [Emphasis added] Id. at 1324, 63 USPQ2d at 1613 .

The court in Enzo adopted its standard from the USPTO's Written Description Examination Guidelines. See 296 F.3d at 1324, 63 USPQ2d at 1613 (citing the Guidelines). The Guidelines apply to proteins as well as DNAs.

Finally, it is well-settled that the written description requirement of 35 U.S.C. § 112, first paragraph, can be satisfied without express or explicit disclosure of a later-claimed invention. See, e.g., In re Herschler, 591 F.2d 693, 700, 200 USPQ 711, 717 (CCPA 1979): "The claimed subject matter need not be described in haec verba to satisfy the description requirement. It is not necessary that the application describe the claim limitations exactly, but only so clearly that one having ordinary skill in the pertinent art would recognize from the disclosure that appellants invented processes including

those limitations." (citations omitted). See also Purdue Pharma L.P. v. Faulding, Inc., 230 F.3d 1320, 1323, 56 USPQ2d 1481, 1483 (Fed. Cir. 2000) ("In order to satisfy the written description requirement, the disclosure as originally filed does not have to provide in haec verba support for the claimed subject matter at issue.").

We apply the relevant law above to the facts before us. In the present case, the examiner argues that the "specification does not set forth what specific structural or physical features define the claimed isolated nucleic acids and transgenic cells, plants and seeds." Answer, page 4. The examiner argues that one skilled in the art "could not predict the structure and function of isolated nucleic acids comprising a wee1 polynucleotide having at least 80% identity to the entire coding region of SEQ ID NO:1 or a polynucleotide complementary thereto, or cells, plants and seeds transformed therewith. The physical features of the claimed isolated nucleic acids and transgenic cells, plants, and seeds cannot be ascertained in the absence of information about the functional activities of these nucleic acids. Additionally, the specification does not disclose the effect of incorporating the claimed isolated nucleic acids into the genome of a cell or plant." Id.

We find the examiner's argument that one skilled in the art could not predict the structure and function of isolated nucleic acids comprising a wee1 to be confusing in the context of a written description rejection, as predictability is not the legal standard or test for such rejections. However, as best we can understand the examiner's argument, the examiner appears to argue that the specification does not describe a wee1

polynucleotide having at least 80% identity to the entire coding region of SEQ ID NO:1.

The examiner argues that "Applicant's [sic] own specification fails to teach a single representative species with 80% identity and WEE1 function." Answer, page 5.

We do not agree with the examiner that claim 31 lacks written description in the specification and that appellants were not in possession of the claimed invention at the time the application was filed. First, to satisfy the written description requirement it is not necessary that the application describe the claim limitations exactly, but only so clearly that one having ordinary skill in the pertinent art would recognize from the disclosure that appellants invented the claimed subject matter. Thus, we do not find the fact that the specification does not specifically teach the structure of a species with 80% identity and WEE1 function to be dispositive of the written description issue here.

The Enzo court stated that "the written description requirement can be met by 'show[ing] that an invention is complete by disclosure of sufficiently detailed, relevant identifying characteristics . . . i.e., complete or partial structure, other physical and/or chemical properties, functional characteristics when coupled with a known or disclosed correlation between function and structure, or some combination of such characteristics.'" Id. at 1324, 63 USPQ2d at 1613 (emphasis omitted, bracketed material in original).

The specification specifically describes the chemical structures of a polynucleotide that encodes a polypeptide of SEQ ID NO:2 and a polynucleotide comprising the coding sequence set forth in SEQ ID NO:1. The specification also provides an example of how to screen for WEE1 activity, specification, Example 1, pages 33-34 and Example 3. Contrary to the examiner's position, it would reasonably appear that such a description in the specification would constitute sufficiently detailed, relevant identifying characteristics of the claimed subject matter consistent with Enzo (*supra*).

In our view, the examiner has failed to indicate why one of ordinary skill in the art, who is in possession of the very specific chemical structures of a polynucleotide that encodes a polypeptide of SEQ ID NO:2 and a polynucleotide comprising the coding sequence set forth in SEQ ID NO:1, would be unable to recognize, upon reading the disclosure, that appellants invented the claimed subject matter, including homologues sharing structural features with the specifically claimed and disclosed structures.

The examiner relies on Aligue for the teaching that amino acids 363-408 of the 550 amino acid N-terminal regulatory domain of *S. pombe* WEE1 are critical to the function of the regulatory domain. The examiner concludes that because "the functional properties of WEE1 and other proteins reside in specific amino acid residues, changes in these residues could have an effect on WEE1 function." Answer, page 5.

We agree with appellants that the examiner has not established with a preponderance of the evidence, that the combination of the disclosure of the specific chemical structures of a polynucleotide comprising the coding sequence set forth in SEQ ID NO:1, as well as teachings in the specification on how to test for wee1 activity and teachings of the areas of the wee1 gene that can be altered without disturbing substrate recognition are insufficient to describe a wee1 polynucleotide having at least 80% identity to the entire coding region of SEQ ID NO:1. What is evident from the record is those of ordinary skill in the art were aware that most of the variations in amino acid sequences of WEE1 are in the amino terminus, while the carboxy end of the genes are relatively conserved. Those of skill in the art were also aware that the carboxyl terminus and the central portion of the WEE1 protein from *S. pombe* contain the protein kinase domains and sequence crucial for substrate recognition and catalysis. Thus, those of ordinary skill in the art would have recognized from reading the disclosure that the inventors had invented the isolated wee1 having the specific nucleotide and amino acid sequences and variations of these sequences with mutations in described specific areas of Wee1, while avoiding the introduction of mutations in other regions. This teaching, coupled with the ability to test for functional mutants with the assays provided for in the specification, supports appellants' position that the inventors sufficiently described and were in possession of the invention as claimed, at the time of filing of the patent application.

In our view the examiner has not provided sufficient evidence or analysis to indicate why one of ordinary skill in the art having read the disclosure, would not have been able to recognize that the inventors invented the subject matter within the scope of the claims. The rejection of the claims for lack of written description is reversed.

Enablement

Claims 2-11, 31, 33 and 35-36 stand rejected under 35 U.S.C. § 112, first paragraph for lack of enablement.

It is the examiner's position that the specification is enabling for an isolated wee1 nucleic acid comprising a polynucleotide encoding SEQ ID NO:2 and a polynucleotide comprising SEQ ID NO:1, but does not reasonably provide enablement for a wee1 polynucleotide having 80% identity to the coding region of SEQ ID NO:1. Answer, page 6.

Enablement is a legal determination of whether a patent enables one skilled in the art to make and use the claimed invention, Raytheon Co. v. Roper Corp., 724 F.2d 951, 960, 220 USPQ 592, 599 (Fed. Cir. 1983), and is not precluded even if some experimentation is necessary, although the amount of experimentation needed must not be unduly extensive. Atlas Powder Co. v. E.I. Du Pont De Nemours & Co., 750 F.2d 1569, 1576, 224 USPQ 409, 413 (Fed. Cir. 1984); W.L. Gore and Associates v. Garlock, Inc., 721 F.2d 1540, 1556, 220 USPQ 303, 315 (Fed. Cir. 1983). Nothing more than objective enablement is required, and therefore it is irrelevant

whether this teaching is provided through broad terminology or illustrative examples. In re Marzocchi, 439 F.2d 220, 223, 169 USPQ 367, 369 (CCPA 1971).

An analysis of whether the claims under appeal are supported by an enabling disclosure requires a determination of whether that disclosure contained sufficient information regarding the subject matter of the appealed claims as to enable one skilled in the pertinent art to make and use the claimed invention. In order to establish a prima facie case of lack of enablement, the examiner has the initial burden to establish a reasonable basis to question the enablement provided for the claimed invention. See In re Wright, 999 F.2d 1557, 1561-62, 27 USPQ2d 1510, 1513 (Fed. Cir. 1993) (examiner must provide a reasonable explanation as to why the scope of protection provided by a claim is not adequately enabled by the disclosure). See also In re Morehouse, 545 F.2d 162, 192 USPQ 29 (CCPA 1976).

The threshold step in resolving this issue is to determine whether the examiner has met his burden of proof by advancing acceptable reasoning inconsistent with enablement. "Factors to be considered by the examiner in determining whether a disclosure would require undue experimentation have been summarized by the board in Ex parte Forman, [230 USPQ 546, 547 (Bd Pat App Int 1986)]. They include (1) the quantity of experimentation necessary, (2) the amount of direction or guidance presented, (3) the presence or absence of working examples, (4) the nature of the invention, (5) the state of the prior art, (6) the relative skill of those in the art, (7) the predictability or unpredictability of the art, and (8) the breadth of the claims." (footnote

omitted). In re Wands, 858 F.2d 731, 737, 8 USPQ2d 1400, 1404, (Fed. Cir. 1988).

In the present case the examiner provided an analysis of several of the relevant enablement factors on pages 5-9 of the Answer. One of the examiner's primary arguments is that the specification does not disclose any specific structural or functional characteristics of any isolated nucleic acid comprising a polynucleotide having at least 80% identity to the entire coding region of SEQ ID NO:1. Answer, page 7. The examiner also argues that the "specification does not disclose any examples of how to make a transgenic host cell or plant comprising an isolated nucleic acid comprising a polynucleotide having at least 80% identity to the entire coding region of SEQ ID NO:1" or provide "any definitive evidence that introducing any isolated nucleic acid comprising a polynucleotide having at least 80% identity to the entire coding region of SEQ ID NO:1 into a plant will result in an alteration of the plant's phenotype." Id.

The examiner relies on Hemerly to support the position that the transformation of plant material is unpredictable in view of the disclosure. According to the examiner, Hemerly teaches "the transformation of *Arabidopsis* and tobacco plants with isolated nucleic acids encoding wild-type and mutant Cdc2a cell cycle regulatory proteins". Answer, page 8. Transformation of *Arabidopsis* with wild-type Cdc2a and with a Cdc2a mutant designed to accelerate the cell cycle unexpectedly did not affect the development of transgenic plants. The transformation of *Arabidopsis* and tobacco with a Cdc2a mutant designed to arrest the cell cycle did affect the development of transgenic plants as expected. Id.

The examiner concludes (Id., pages 8-9)

> Given the unpredictability of determining the function of isolated nucleic acids comprising a polynucleotide having at least 80% identity to the entire coding region of SEQ ID NO:1, the unpredictability of altering the phenotype of a plant by transforming it with an isolated nucleic acid of SEQ ID NO:1 or isolated nucleic acids comprising a polynucleotide having at least 80% identity to the entire coding region of SEQ ID NO:1, the absence of guidance in the specification for making and using said nucleic acids and transgenic host cells, plants, and seeds, the lack of working examples, and given the breadth of the claims which encompass multiple polynucleotides having at least 80% identity to the entire coding region of SEQ ID NO:1, it would require undue experimentation by one skilled in the art to make and/or use the claimed invention.

Analysis of the enablement requirement in the present case dovetails with our analysis with respect to the written description requirement. In particular, the specification specifically describes the chemical structures of a polynucleotide that encodes a polypeptide of SEQ ID NO:2 and a polynucleotide comprising the coding sequence set forth in SEQ ID NO:1. The specification also provides an example of how to screen for WEE1 activity, specification, Example 1, pages 33-34 and Example 3. Brief, page 9. In addition, the specification page 3, lines 17-31, "describes the level of skill in the art as well as indicating areas of the wee1 gene that can be altered without disturbing substrate recognition." Brief, page 7. Moreover, the specification, page 3, states, "Most of the variations in amino acid sequences of WEE1 are in the amino terminus, while the carboxy end of the genes are relatively conserved. The carboxyl terminus and the central portion of the WEE1 protein from *S. pombe* contain the protein kinase domains and sequence crucial for substrate recognition and catalysis."

Ex parte Jackson, 217 USPQ 804, 807 (1982).

In our view, upon reading the disclosure, those of ordinary skill in the art would have been provided a reasonable amount of guidance to make and use a wee1 polynucleotide having at least 80% identity to the entire coding region of SEQ ID NO:1. The specification, pages 27-29 outlines methods for transfection and transformation of cells and the introduction of DNA into plants. The examples of the specification indicate successful expression of zmwee1 in E. coli as evidenced by the successful inhibition of cyclin-dependent protein kinase. Specification, pages 33-34. In view of the successful transformation of cells with the disclosed and claimed specific wee1, we find no evidence or sufficient indicated reason of record why one of ordinary skill in the art would not have had a reasonable expectation of success in transforming cells and plant cells with a wee1 polynucleotide having at least 80% identity to the entire coding region of SEQ ID NO:1 without undue experimentation.

The rejection of the claims for lack of enablement is reversed.

CONCLUSION

The rejection of claims 2-11, 31, 33 and 35-36 under 35 U.S.C. § 112, first paragraph as containing subject matter that was not described in the specification in such a way as to reasonably convey to one skilled in the art at the time the application was filed that the inventor had possession of the claimed invention is reversed.

We agree with appellants that the examiner has not established that the combination of the disclosure of the specific chemical structures of a polynucleotide comprising the coding sequence set forth in SEQ ID NO:1, as well as teachings in the specification on how to test for wee1 activity and teachings of the areas of the wee1 gene that can be altered without disturbing substrate recognition are insufficient to enable a wee1 polynucleotide having at least 80% identity to the entire coding region of SEQ ID NO:1.

Nor has the examiner established that one of ordinary skill in the art having the chemical structures of a polynucleotide comprising the coding sequence set forth in SEQ ID NO:1 and the ability to test for expression as described in the specification, would be insufficient to transform cells, plants and seeds in view of the success described in the specification. While the examiner relies on Hemerly for the transformation of *Arabidopsis* with wild-type Cdc2a and with a Cdc2a mutant, the examiner has not explained how or why potential unpredictability associated with Cdc2a expression is related to or affects Wee1 expression. Nor is it clear from the examiner's analysis that the examiner has fully considered the state of the art as it relates to the transformation of vectors, seeds and plant cells, as outlined in the specification.

The Patent and Trademark Office Board of Appeals stated:

> The test [for enablement] is not merely quantitative, since a considerable amount of experimentation is permissible, if it is merely routine, or if the specification in question provides a reasonable amount of guidance with respect to the direction in which the experimentation should proceed to enable the determination of how to practice a desired embodiment of the invention claimed.

The rejection of claims 2-11, 31, 33 and 35-36 under 35 U.S.C. § 112, first paragraph for lack of enablement is reversed.

No time period for taking any subsequent action in connection with this appeal may be extended under 37 CFR § 1.136(a).

REVERSED

WILLIAM F. SMITH Administrative Patent Judge	)
DEMETRA J. MILLS Administrative Patent Judge	) APPEALS AND
	) INTERFERENCES
ERIC GRIMES Administrative Patent Judge	)

Appeal No. 2003-1993
Application No. 09/470,526

PIONEER HI-BRED INTERNATIONAL, INC.
7100 N.W. 62nd Ave.
P.O. Box 1000
Johnson, IA 50131

Deciphering the Message in Protein Sequences: Tolerance to Amino Acid Substitutions

JAMES U. BOWIE,* JOHN F. REIDHAAR-OLSON, WENDELL A. LIM, ROBERT T. SAUER

An amino acid sequence encodes a message that determines the shape and function of a protein. This message is highly degenerate in that many different sequences can code for proteins with essentially the same structure and activity. Comparison of different sequences with similar messages can reveal key features of the code and improve understanding of how a protein folds and how it performs its function.

THE GENOME IS MANIFEST LARGELY IN THE SET OF PROteins that it encodes. It is the ability of these proteins to fold into unique three-dimensional structures that allows them to function and carry out the instructions of the genome. Thus, comprehending the rules that relate amino acid sequence to structure is fundamental to an understanding of biological processes. Because an amino acid sequence contains all of the information necessary to determine the structure of a protein (*1*), it should be possible to predict structure from sequence, and subsequently to infer detailed aspects of function from the structure. However, both problems are extremely complex, and it seems unlikely that either will be solved in an exact manner in the near future. It may be possible to obtain approximate solutions by using experimental data to simplify the problem. In this article, we describe how an analysis of allowed amino acid substitutions in proteins can be used to reduce the complexity of sequences and reveal important aspects of structure and function.

Methods for Studying Tolerance to Sequence Variation

There are two main approaches to studying the tolerance of an amino acid sequence to change. The first method relies on the process of evolution, in which mutations are either accepted or rejected by natural selection. This method has been extremely powerful for proteins such as the globins or cytochromes, for which sequences from many different species are known (*2–7*). The second approach uses genetic methods to introduce amino acid changes at specific positions in a cloned gene and uses selections or screens to identify functional sequences. This approach has been used to great advantage for proteins that can be expressed in bacteria or yeast, where the appropriate genetic manipulations are possible (*3, 8–11*). The end results of both methods are lists of active sequences that can be compared and analyzed to identify sequence features that are essential for folding or function. If a particular property of a side chain, such as charge or size, is important at a given position, only side chains that have the required property will be allowed. Conversely, if the chemical identity of the side chain is unimportant, then many different substitutions will be permitted.

Studies in which these methods were used have revealed that proteins are surprisingly tolerant of amino acid substitutions (*2–4, 11*). For example, in studying the effects of approximately 1500 single amino acid substitutions at 142 positions in *lac* repressor, Miller and co-workers found that about one-half of all substitutions were phenotypically silent (*11*). At some positions, many different, nonconservative substitutions were allowed. Such residue positions play little or no role in structure and function. At other positions, no substitutions or only conservative substitutions were allowed. These residues are the most important for *lac* repressor activity.

What roles do invariant and conserved side chains play in proteins? Residues that are directly involved in protein functions such as binding or catalysis will certainly be among the most conserved. For example, replacing the Asp in the catalytic triad of trypsin with Asn results in a 10^4-fold reduction in activity (*12*). A similar loss of activity occurs in λ repressor when a DNA binding residue is changed from Asn to Asp (*13*). To carry out their function, however, these catalytic residues and binding residues must be precisely oriented in three dimensions. Consequently, mutations in residues that are required for structure formation or stability can also have dramatic effects on activity (*10, 14–16*). Hence, many of the residues that are conserved in sets of related sequences play structural roles.

Substitutions at Surface and Buried Positions

In their initial comparisons of the globin sequences, Perutz and co-workers found that most buried residues require nonpolar side chains, whereas few features of surface side chains are generally conserved (*6*). Similar results have been seen for a number of protein families (*2, 4, 5, 7, 17, 18*). An example of the sequence tolerance at surface versus buried sites can be seen in Fig. 1, which shows the allowed substitutions in λ repressor at residue positions that are near the dimer interface but distant from the DNA binding surface of the protein (*9*). These substitutions were identified by a functional

The authors are in the Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139.

*Present address: Department of Chemistry and Biochemistry and the Molecular Biology Institute, University of California, Los Angeles, Los Angeles, CA 90024.

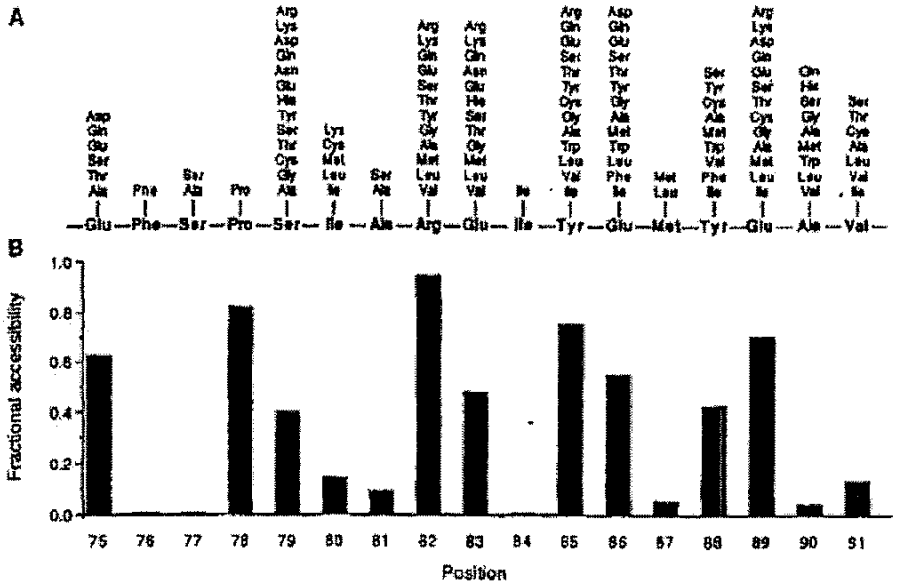


Fig. 1. (A) Amino acid substitutions allowed in a short region of λ repressor. The wild-type sequence is shown along the center line. The allowed substitutions shown above each position were identified by randomly mutating one to three codons at a time by using a cassette method and applying a functional selection (*9*). (B) The fractional solvent accessibility (*42*) of the wild-type side chain in the protein dimer (*43*) relative to the same atoms in an Ala-X-Ala model tripeptide.

selection after cassette mutagenesis. A histogram of side chain solvent accessibility in the crystal structure of the dimer is also shown in Fig. 1. At six positions, only the wild-type residue or relatively conservative substitutions are allowed. Five of these positions are buried in the protein. In contrast, most of the highly exposed positions tolerate a wide range of chemically different side chains, including hydrophilic and hydrophobic residues. Hence, it seems that most of the structural information in this region of the protein is carried by the residues that are solvent inaccessible.

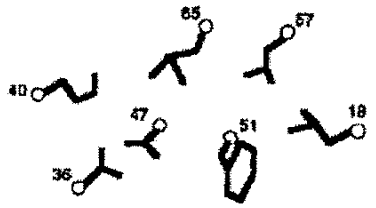
Constraints on Core Sequences

Because core residue positions appear to be extremely important for protein folding or stability, we must understand the factors that dictate whether a given core sequence will be acceptable. In general, only hydrophobic or neutral residues are tolerated at buried sites in proteins, undoubtedly because of the large favorable contribution of the hydrophobic effect to protein stability (*19*). For example, Fig. 2 shows the results of genetic studies used to investigate the substitutions allowed at residue positions that form the hydrophobic core of the NH_2-terminal domain of λ repressor (*20*). The acceptable core sequences are composed almost exclusively of Ala, Cys, Thr, Val, Ile, Leu, Met, and Phe. The acceptability of many different residues at each core position presumably reflects the fact that the hydrophobic effect, unlike hydrogen bonding, does not depend on specific residue pairings. Although it is possible to imagine a hypothetical core structure that is stabilized exclusively by residues forming hydrogen bonds and salt bridges, such a core would probably be difficult to construct because hydrogen bonds require pairing of donors and acceptors in an exact geometry. Thus the repertoire of possible structures that use a polar core would probably be extremely limited (*21*). Polar and charged residues are occasionally found in the cores of proteins, but only at positions where their hydrogen bonding needs can be satisfied (*22*).

The cores of most proteins are quite closely packed (*23*), but some volume changes are acceptable. In λ repressor, the overall core volume of acceptable sequences can vary by about 10%. Changes at individual sites, however, can be considerably larger. For example, as shown in Fig. 2, both Phe and Ala are allowed at the same core position in the appropriate sequence contexts. Large volume changes at individual buried sites have also been observed in phylogenetic studies, where it has been noted that the size decreases and increases at interacting residues are not necessarily related in a simple complementary fashion (*5*, *7*, *17*). Rather, local volume changes are accommodated by conformational changes in nearby side chains and by a variety of backbone movements.

The Informational Importance of the Core

With occasional exceptions, the core must remain hydrophobic and maintain a reasonable packing density. However, since the core is composed of side chains that can assume only a limited number of conformations (*24*), efficient packing must be maintained without steric clashes. How important are hydrophobicity, volume, and steric complementarity in determining whether a given sequence can form an acceptable core? Each factor is essential in a physical sense, as a stable core is probably unable to tolerate unsatisfied hydrogen bonding groups, large holes, or steric overlaps (*25*). However, in an informational sense, these factors are not equivalent. For example, in experiments in which three core residues of λ repressor were mutated simultaneously, volume was a relatively unimportant informational constraint because three-quarters of all possible combinations of the 20 naturally occurring amino acids had volumes within the range tolerated in the core, and yet most of these sequences were unacceptable (*20*). In contrast, of the sequences that contained only



40	36	47	65	51	57	18
Ala	Ala	Ala	Ala	Cys	Cys	Ala
Cys	Cys	Cys	Ser	Val	Pro	Cys
Val	Thr	Thr	Cys	Ile	Val	Val
Ile	Val	Val	Thr	Leu	Ile	Ile
Leu	Ile	Ile	Val	Met	Leu	Leu
Met	Leu	Leu	Ile	Phe	Met	Met
Phe	Met	Met	Leu		Phe	
	Phe	Phe	Met			
			Phe			

Fig. 2. Amino acid substitutions allowed in the core of λ repressor. The wild-type side chains are shown pictorially in the approximate orientation seen in the crystal structure (*43*). The lists of allowed substitutions at each position are shown below the wild-type side chains. These substitutions were identified by randomly mutating one to four residues at a time by using a cassette method and applying a functional selection (*20*). Not all substitutions are allowed in every sequence background.

the appropriate hydrophobic residues, a significant fraction were acceptable. Hence, the hydrophobicity of a sequence contains more information about its potential acceptability in the core than does the total side chain volume. Steric compatibility was intermediate between volume and hydrophobicity in informational importance.

The Informational Importance of Surface Sites

We have noted that many surface sites can tolerate a wide variety of side chains, including hydrophilic and hydrophobic residues. This result might be taken to indicate that surface positions contain little structural information. However, Bashford *et al.*, in an extensive analysis of globin sequences (*4*), found a strong bias against large hydrophobic residues at many surface positions. At one level, this may reflect constraints imposed by protein solubility, because large patches of hydrophobic surface residues would presumably lead to aggregation. At a more fundamental level, protein folding requires a partitioning between surface and buried positions. Consequently, to achieve a unique native state without significant competition from other conformations, it may be important that some sites have a decided preference for exterior rather than interior positions. As a result, many surface sites can accept hydrophobic residues individually, but the surface as a whole can probably tolerate only a moderate number of hydrophobic side chains.

Identification of Residue Roles from Sets of Sequences

Often, a protein of interest is a member of a family of related sequences. What can we infer from the pattern of allowed substitutions at positions in sets of aligned sequences generated by genetic or phylogenetic methods? Residue positions that can accept a number of different side chains, including charged and highly polar residues, are almost certain to be on the protein surface. Residue positions that remain hydrophobic, whether variable or not, are likely to be buried within the structure. In Fig. 3, those residue positions in λ repressor that can accept hydrophilic side chains are shown in orange and those that cannot accept hydrophilic side chains are shown in green. The obligate hydrophobic positions define the core of the structure, whereas positions that can accept hydrophilic side chains define the surface.

Functionally important residues should be conserved in sets of active sequences, but it is not possible to decide whether a side chain is functionally or structurally important just because it is invariant or conserved. To make this distinction requires an independent assay of protein folding. The ability of a mutant protein to maintain a stably folded structure can often be measured by biophysical techniques, by susceptibility to intracellular proteolysis (*26*), or by binding to antibodies specific for the native structure (*27, 28*). In the latter cases, it is possible to screen proteins in mutated clones for the ability to fold even if these proteins are inactive. Sets of sequences that allow formation of a stable structure can then be compared to the sets that allow both folding and function, with the active site or binding residues being those that are variable in the set of stable proteins but invariant in the set of functional proteins. The DNA-binding residues of Arc repressor were identified by this method (*8*). The receptor-binding residues of human growth hormone were also identified by comparing the stabilities and activities of a set of mutant sequences (*28*). However, in this case, the mutants were generated as hybrid sequences between growth hormone and related hormones with different binding specificities.

Implications for Structure Prediction

At present, the only reliable method for predicting a low-resolution tertiary structure of a new protein is by identifying sequence similarity to a protein whose structure is already known (*29, 30*). However, it is often difficult to align sequences as the level of sequence similarity decreases, and it is sometimes impossible to detect statistically significant sequence similarity between distantly related proteins. Because the number of known sequences is far greater than the number of known structures, it would be advantageous to increase the reach of the available structural information by improving methods for detecting distant sequence relations and for subsequently aligning these sequences based on structural principles. In a normal homology search, the sequence database is scanned with a single test sequence, and every residue must be weighted equally. However, some residues are more important than others and should be weighted accordingly. Moreover, certain regions of the protein are more likely to contain gaps than others. Both kinds of information can be obtained from sequence sets, and several techniques have

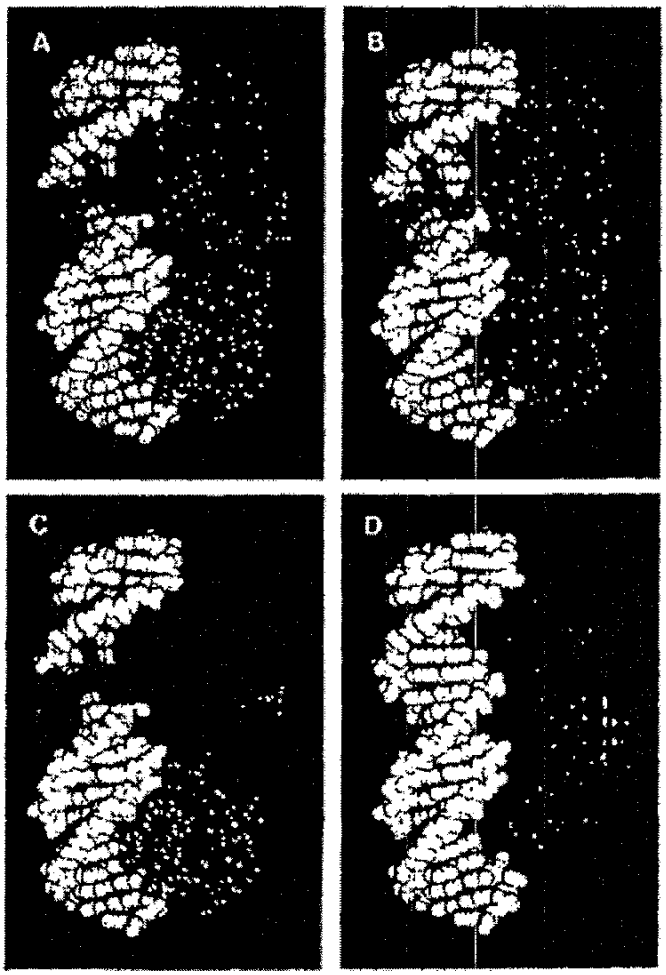


Fig. 3. Tolerance of positions in the NH_2-terminal domain of λ repressor to hydrophilic side chains. The complex (*43*) of the repressor dimer (blue) and operator DNA (white) is shown. In (**A**), positions that can tolerate hydrophilic side chains are shown in orange. The same side chains are shown in (**B**) without the remaining protein atoms. In (**C**), positions that require hydrophobic or neutral side chains are shown in green. These side chains are shown in (**D**) without the remaining protein atoms. About three-fourths of the 92 side chains in the NH_2-terminal domain are included in both (**B**) and (**D**). The remaining positions have not been tested. Data are from (*9, 14, 40, 41, 44*).

been used to combine such information into more appropriately weighted sequence searches and alignments (*31*). These methods were used to align the sequences of retroviral proteases with aspartic proteases, which in turn allowed construction of a three-dimensional model for the protease of human immunodeficiency virus type 1 (*29*). Comparison with the recently determined crystal structure of this protein revealed reasonable agreement in many areas of the predicted structure (*32*).

The structural information at most surface sites is highly degenerate. Except for functionally important residues, exterior positions seem to be important chiefly in maintaining a reasonably polar surface. The information contained in buried residues is also degenerate, the main requirement being that these residues remain hydrophobic. Thus, at its most basic level, the key structural message in an amino acid sequence may reside in its specific pattern of hydrophobic and hydrophilic residues. This is meant in an informational sense. Clearly, the precise structure and stability of a protein depends on a large number of detailed interactions. It is possible, however, that structural prediction at a more primitive level can be accomplished by concentrating on the most basic informational aspects of an amino acid sequence. For example, amphipathic patterns can be extracted from aligned sets of sequences and used, in some cases, to identify secondary structures.

If a region of secondary structure is packed against the hydrophobic core, a pattern of hydrophobic residues reflecting the periodicity of the secondary structure is expected (*33, 34*). These patterns can be obscured in individual sequences by hydrophobic residues on the protein surface. It is rare, however, for a surface position to remain hydrophobic over the course of evolution. Consequently, the amphipathic patterns expected for simple secondary structures can be much clearer in a set of related sequences (*6*). This principle is illustrated in Fig. 4, which shows helical hydrophobic moment plots for the Antennapedia homeodomain sequence (Fig. 4A) and for a composite sequence derived from a set of homologous homeodomain proteins (Fig. 4B) (*35*). The hydrophobic moment is a simple measure of the degree of amphipathic character of a sequence in a given secondary structure (*34*). The amphipathic character of the three α-helical regions in the Antennapedia protein (*36*) is clearly revealed only by the analysis of the combined set of homeodomain sequences. The secondary structure of Arc repressor, a small DNA-binding protein, was recently predicted by a similar method (*8*) and confirmed by nuclear magnetic resonance studies (*37*).

The specific pattern of hydrophobic and hydrophilic residues in an amino acid sequence must limit the number of different structures a given sequence can adopt and may indeed define its overall fold. If this is true, then the arrangement of hydrophobic and hydrophilic residues should be a characteristic feature of a particular fold. Sweet and Eisenberg have shown that the correlation of the pattern of hydrophobicity between two protein sequences is a good criterion for their structural relatedness (*38*). In addition, several studies indicate that patterns of obligatory hydrophobic positions identified from aligned sequences are distinctive features of sequences that adopt the same structure (*4, 29, 38, 39*). Thus, the order of hydrophobic and hydrophilic residues in a sequence may actually be sufficient information to determine the basic folding pattern of a protein sequence.

Although the pattern of sequence hydrophobicity may be a characteristic feature of a particular fold, it is not yet clear how such patterns could be used for prediction of structure de novo. It is important to understand how patterns in sequence space can be related to structures in conformation space. Lau and Dill have approached this problem by studying the properties of simple sequences composed only of H (*hydrophobic*) and P (*polar*) groups on two-dimensional lattices (*40*). An example of such a representation is shown in Fig. 5. Residues adjacent in the sequence must occupy adjacent squares on the lattice, and two residues cannot occupy the same space. Free energies of particular conformations are evaluated with a single term, an attraction of H groups. By considering chains of ten residues, an exhaustive conformational search for all 1024 possible sequences of H and P residues was possible. For longer sequences only a representative fraction of the allowed sequence or conformation space could be explored. The significant results were as follows: (i) not all sequences can fold into a "native" structure and only a few sequences form a unique native structure; (ii) the probability that a sequence will adopt a unique native structure increases with chain length; and (iii) the native states are compact, contain a hydrophobic core surrounded by polar residues, and contain significant secondary structure. Although the gap between these two-dimensional simulations and three-dimensional structures is large, the use of simple rules and sequence representations yields results similar to those expected for real proteins. Three-dimensional lattice methods are also beginning to be developed and evaluated (*41*).

Summary

There is more information in a set of related sequences than in a single sequence. A number of practical applications arise from an analysis of the tolerance of residue positions to change. First, such information permits the evaluation of a residue's importance to the function and stability of a protein. This ability to identify the essential elements of a protein sequence may improve our understanding of the determinants of protein folding and stability as well as protein function. Second, patterns of tolerance to amino acid substitutions of varying hydrophilicity can help to identify residues likely to be buried in a protein structure and those likely to occupy

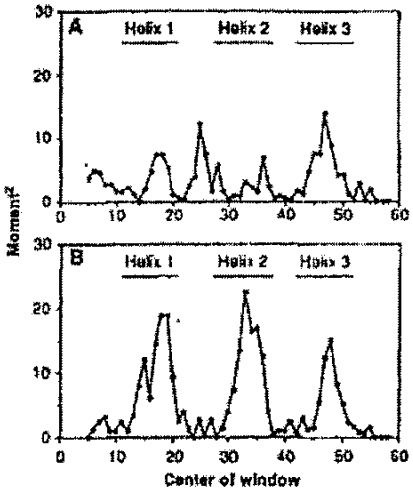


Fig. 4. Helical hydrophobic moments calculated by using (**A**) the Antennapedia homeodomain sequence or (**B**) a set of 39 aligned homeodomain sequences (*35*). The bars indicate the extent of the helical regions identified in nuclear magnetic resonance studies of the Antennapedia homeodomain (*36*). To determine hydrophobic moments, residues were assigned to one of three groups: H1 (high hydrophobicity = Trp, Ile, Phe, Leu, Met, Val, or Cys), H2 (medium hydrophobicity = Tyr, Pro, Ala, Thr, His, Gly, or Ser); and H3 (low hydrophobicity = Gln, Asn, Glu, Asp, Lys, or Arg). For the aligned homeodomain sequences, the residues at each position were sorted by their hydrophobicity by using the scale of Fauchere and Pliska (*45*). Arg and Lys were not counted unless no other residue was found at the position, because they contain long aliphatic side chains and can thereby substitute for nonpolar residues at some buried sites. To account for possible sequence errors and rare exceptions, the most hydrophilic residue allowed at each position was discarded unless it was observed twice. The second most hydrophilic residue was then chosen to represent the hydrophobicity of each position. An eight-residue window was used and the vectors projected radially every 100°. The vector magnitudes were assigned a value of 1, 0, or −1 for positions where the hydrophobicity group was H1, H2, or H3, respectively.

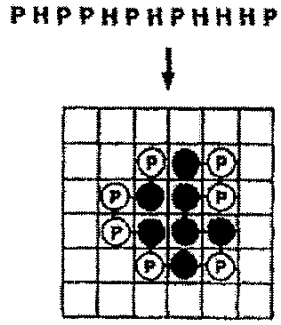


Fig. 5. A representation of one compact conformation for a particular sequence of H and P residues on a two-dimensional square lattice. [Adapted from (*40*), with permission of the American Chemical Society]

surface positions. The amphipathic patterns that emerge can be used to identify probable regions of secondary structure. Third, incorporating a knowledge of allowed substitutions can improve the ability to detect and align distantly related proteins because the essential residues can be given prominence in the alignment scoring.

As more sequences are determined, it becomes increasingly likely that a protein of interest is a member of a family of related sequences. If this is not the case, it is now possible to use genetic methods to generate lists of allowed amino acid substitutions. Consequently, at least in the short term, it may not be necessary to solve the folding problem for individual protein sequences. Instead, information from sequence sets could be used. Perhaps by simplifying sequence space through the identification of key residues, and by simplifying conformation space as in the lattice methods, it will be possible to develop algorithms to generate a limited number of trial structures. These trial structures could then, in turn, be evaluated by further experiments and more sophisticated energy calculations.

REFERENCES AND NOTES

1. C. J. Epstein, R. F. Goldberger, C. B. Anfinsen, *Cold Spring Harbor Symp. Quant. Biol.* **28**, 439 (1963); C. B. Anfinsen, *Science* **181**, 223 (1973).
2. R. E. Dickerson, *Sci. Am.* **242**, 136 (March 1980).
3. M. D. Hampsey, G. Das, F. Sherman, *FEBS Lett.* **231**, 275 (1988).
4. D. Bashford, C. Chothia, A. M. Lesk, *J. Mol. Biol.* **196**, 199 (1987).
5. A. M. Lesk and C. Chothia, *ibid.* **136**, 225 (1980).
6. M. F. Perutz, J. C. Kendrew, H. C. Watson, *ibid.* **13**, 669 (1965).
7. C. Chothia and A. M. Lesk, *Cold Spring Harbor Symp. Quant. Biol.* **52**, 399 (1965).
8. J. U. Bowie and R. T. Sauer, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 2152 (1989).
9. J. F. Reidhaar-Olson and R. T. Sauer, *Science* **241**, 53 (1988); *Proteins Struct. Funct. Genet.*, in press.
10. D. Shortle, *J. Biol. Chem.* **264**, 5315 (1989).
11. J. H. Miller *et al.*, *J. Mol. Biol.* **131**, 191 (1979).
12. S. Sprang *et al.*, *Science* **237**, 905 (1987); C. S. Craik, S. Rocniak, C. Largman, W. J. Rutter, *ibid.*, p. 909.
13. H. C. M. Nelson and R. T. Sauer, *J. Mol. Biol.* **192**, 27 (1986).
14. M. H. Hecht, J. M. Sturtevant, R. T. Sauer, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 5685 (1984).
15. T. Alber, D. Sun, J. A. Nye, D. C. Muchmore, B. W. Matthews, *Biochemistry* **26**, 3754 (1987).
16. D. Shortle and A. K. Meeker, *Proteins Struct. Funct. Genet.* **1**, 81 (1986).
17. A. M. Lesk and C. Chothia, *J. Mol. Biol.* **160**, 325 (1982).
18. W. R. Taylor, *ibid.* **188**, 233 (1986).
19. W. Kauzmann, *Adv. Protein Chem.* **14**, 1 (1959); R. L. Baldwin, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 8069 (1986).
20. W. A. Lim and R. T. Sauer, *Nature* **339**, 31 (1989); in preparation.
21. Lesk and Chothia (*5*) have argued that a protein core composed solely of hydrogen bonded residues would also be unviable on evolutionary grounds, as a mutational change in one core residue would require compensating changes in any interacting residue or residues to maintain a stable structure.
22. T. M. Gray and B. W. Matthews, *J. Mol. Biol.* **175**, 75 (1984); E. N. Baker and R. E. Hubbard, *Prog. Biophys. Mol. Biol.* **44**, 97 (1984).
23. F. M. Richards, *J. Mol. Biol.* **82**, 1 (1974).
24. J. W. Ponder and F. M. Richards, *ibid.* **193**, 775 (1987).
25. J. T. Kellis, Jr., K. Nyberg, A. R. Fersht, *Biochemistry* **28**, 4914 (1989); W. S. Sandberg and T. C. Terwilliger, *Science* **245**, 54 (1989).
26. A. A. Pakula and R. T. Sauer, *Proteins Struct. Funct. Genet.* **5**, 202 (1989).
27. B. C. Cunningham and J. A. Wells, *Science* **244**, 1081 (1989); R. M. Breyer and R. T. Sauer, *J. Biol. Chem.* **264**, 13348 (1989).
28. B. C. Cunningham, P. Jhurani, P. Ng, J. A. Wells, *Science* **243**, 1330 (1989).
29. L. H. Pearl and W. R. Taylor, *Nature* **329**, 351 (1987).
30. W. J. Brown *et al.*, *J. Mol. Biol.* **42**, 65 (1969); J. Greer, *ibid.* **153**, 1027 (1981); J. M. Berg, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 99 (1988).
31. W. R. Taylor, *Protein Eng.* **2**, 77 (1988).
32. M. A. Navia *et al.*, *Nature* **337**, 615 (1989).
33. M. Schiffer and A. B. Edmundson, *Biophys. J.* **7**, 121 (1967); V. I. Lim, *J. Mol Biol.* **88**, 857 (1974); *ibid.*, p. 873.
34. D. Eisenberg, R. M. Weiss, T. C. Terwilliger, *Nature* **299**, 371 (1982); D. Eisenberg, D. Schwarz, M. Komaromy, R. Wall, *J. Mol. Biol.* **179**, 125 (1984); D. Eisenberg, R. M. Weiss, T. C. Terwilliger, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 140 (1984).
35. T. R. Burglin, *Cell* **53**, 339 (1988).
36. G. Orting *et al.*, *EMBO J.* **7**, 4305 (1988).
37. I. N. Breg, R. Boelens, A. V. E. George, R. Kaptein, *Biochemistry* **28**, 9826 (1989); M. G. Zagorski, J. U. Bowie, A. K. Vershon, R. T. Sauer, D. J. Patel, *ibid.*, p. 9813.
38. R. M. Sweet and D. Eisenberg, *J. Mol. Biol.* **171**, 479 (1983).
39. J. U. Bowie, N. D. Clarke, C. O. Pabo, R. T. Sauer, *Proteins Struct. Funct., Genet.*, in preparation.
40. K. F. Lau and K. A. Dill, *Macromolecules* **22**, 3986 (1989).
41. A. Sikorski and J. Skolnick, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 2668 (1989); A. Kolinski, J. Skolnick, R. Yaris, *Biopolymers* **26**, 937 (1987); D. G. Covell and R. L. Jernigan, *Biochemistry*, in press.
42. B. Lee and F. M. Richards, *J. Mol. Biol.* **55**, 379 (1971).
43. S. R. Jordan and C. O. Pabo, *Science* **242**, 893 (1988).
44. R. M. Breyer, thesis, Massachusetts Institute of Technology, Cambridge (1988).
45. J.-L. Fauchere and V. Pliska, *Eur. J. Med. Chem.-Chim. Ther.* **18**, 369 (1983).

46. We thank C. O. Pabo and S. Jordan for coordinates of the NH_2-terminal domain of λ repressor and its operator complex. We also thank P. Schimmel for the use of his graphics system and J. Bumbaum and C. Francklyn for assistance. Supported in part by NIH grant AI-15706 and predoctoral grants from NSF (J.R.-O.) and Howard Hughes Medical Institute (W.A.L.).